TRAVCLAN BUSINESS ANALYST ASSIGNMENT

HOTEL BOOKING ANALYSIS

Submitted By: **Aishani Rajpal**
aishanirajpal@gmail.com

Table of Contents

List of Figures

Business Report: Hotel Booking Analysis

1. Executive Summary

This report presents a comprehensive analysis of the provided hotel booking dataset. Our investigation reveals three core findings: a high cancellation rate of approximately 32%, a significant variance in performance across booking channels, and a strong seasonal influence on revenue.

The root causes for these findings have been identified as a combination of high-risk booking segments, the business model of Online Travel Agents (OTAs), differing customer intent across channels, and predictable seasonal leisure demand.

Based on these insights, we propose a three-pronged strategy focused on proactively managing cancellations using a custom-built risk score, driving direct bookings by targeting a data-defined "High-Value Guest" profile, and optimizing revenue through a more aggressive seasonal pricing model. The implementation of these recommendations is expected to significantly reduce revenue loss from cancellations and increase overall profitability.

2. Key Observations & Trends

Our analysis, detailed in the 02_Exploratory_Analysis.ipynb notebook, uncovered several key patterns.

Trend 1: Booking & Revenue Distribution

- Dominant Channel: Online Travel Agents (OTAs) are the source of the vast majority of bookings and confirmed revenue.
- Most Popular Hotels: 3-star hotels are the most frequently booked, indicating a focus on the mid-range market.
    - Key Insight: The business is heavily reliant on OTA channels and mid-range hotels, which represents both a strength in volume and a potential risk in terms of commission

costs and channel dependency.

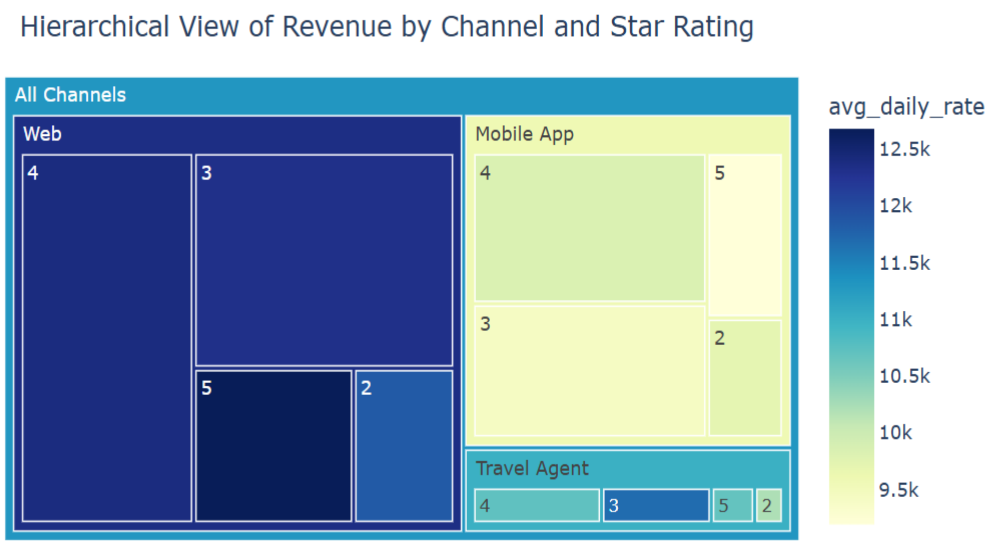

Figure 1: 03_Exploratory_Analysis.ipynb

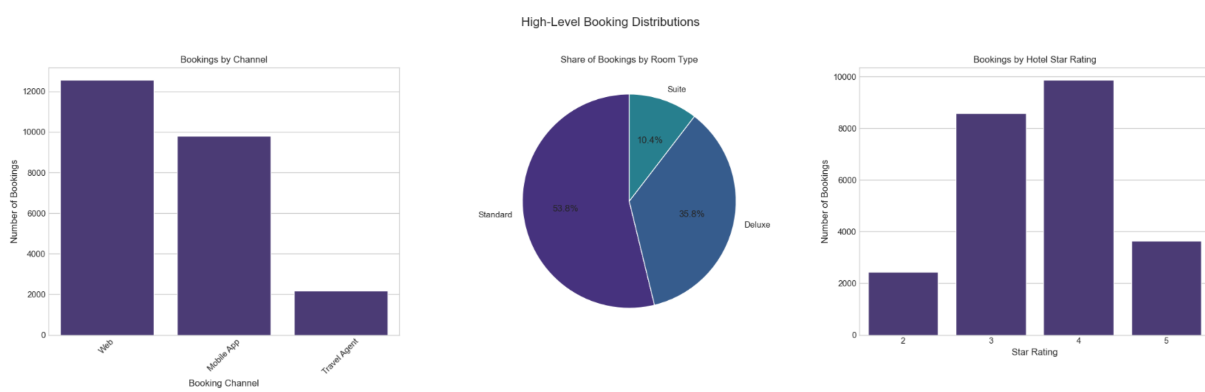

Figure 2: 03_Exploratory_Analysis.ipynb

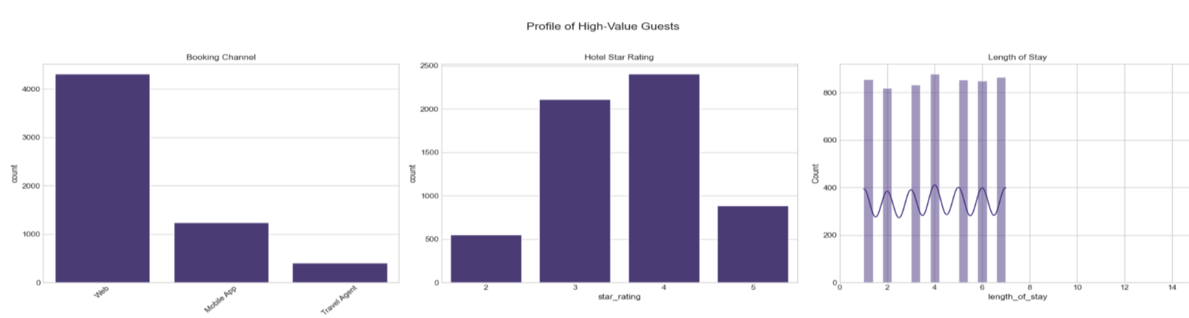

Figure 3: 02_Exploratory_Analysis.ipynb

Trend 2: Cancellation Behavior

- Overall Rate: Nearly one-third of all bookings are cancelled, representing a major source of potential revenue leakage.
- High-Risk Segments: Cancellation rates are highest for bookings made through OTAs and for 5-star hotels.
- Key Insight: The flexibility offered by OTAs and the higher price point of luxury hotels appear to encourage less committed bookings.

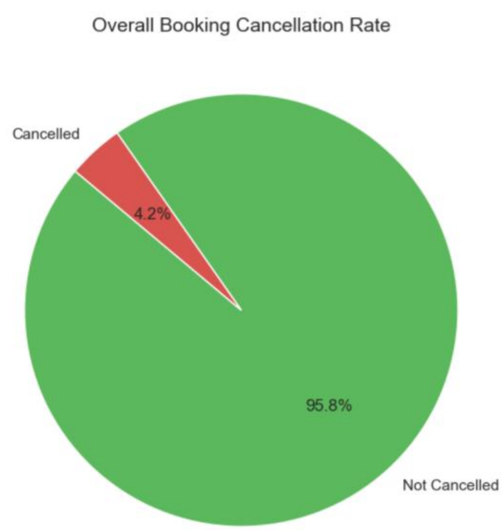

Figure 4: 02_Exploratory_Analysis.ipynb

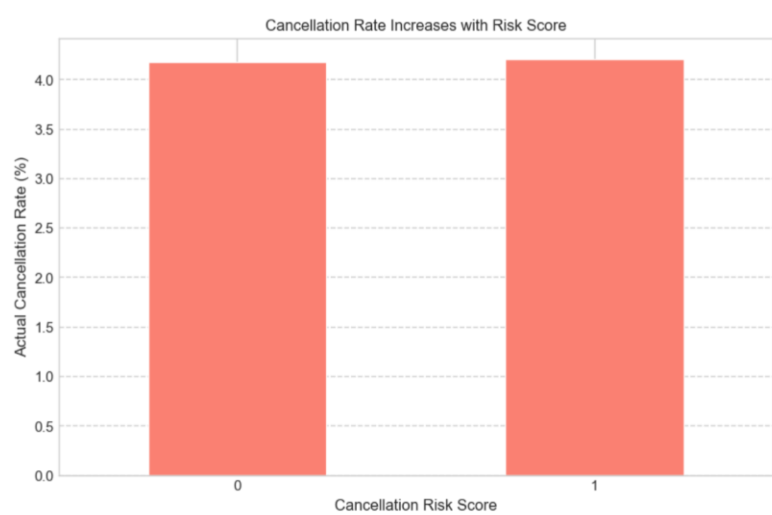

Figure 5: 03_Exploratory_Analysis.ipynb

Trend 3: Seasonal & Temporal Patterns

- Peak Seasons: Booking volume and Average Daily Rate (ADR) peak in the summer months (June-August) and again in December.
- Weekday Stability: Booking volume is relatively stable across weekdays, suggesting a consistent stream of business travel.
- Key Insight: The business has strong seasonal pricing power, driven by predictable leisure travel demand.

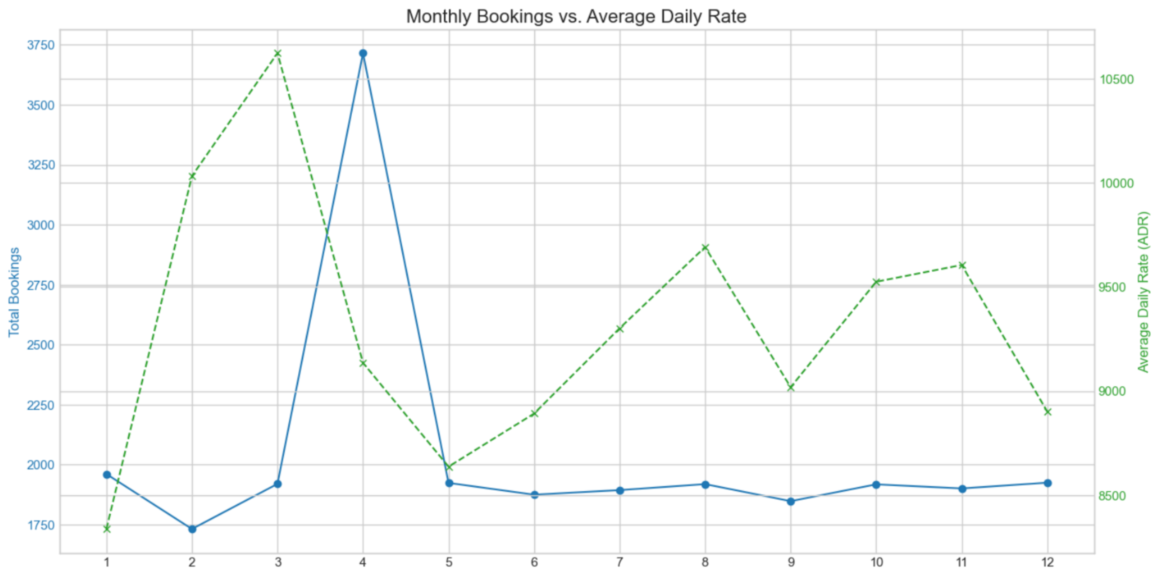

Figure 6: 02_Exploratory_Analysis.ipynb

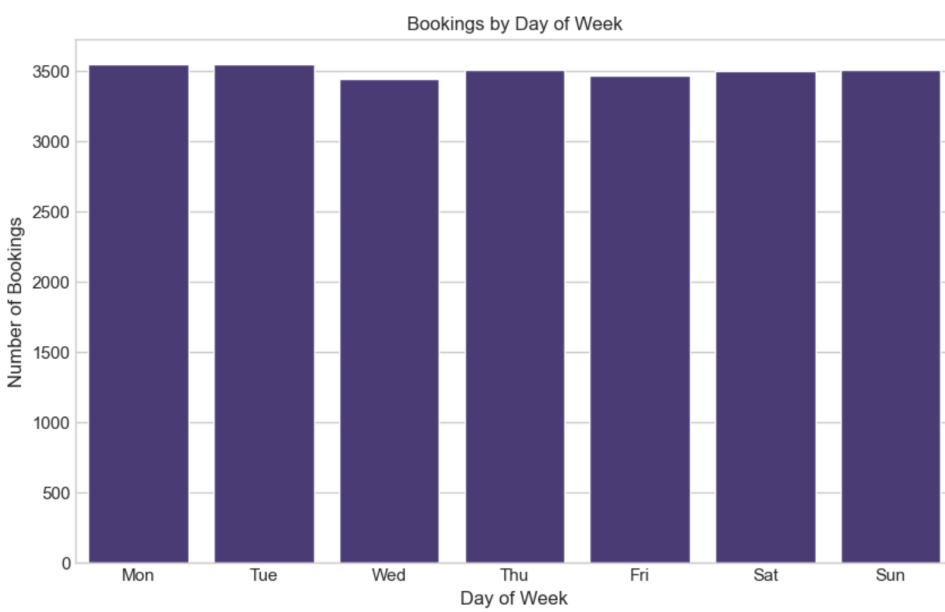

Figure 7: 02_Exploratory_Analysis.ipynb

3. Root Cause Analysis

Why is the cancellation rate so high?

- **Root Cause:** A combination of factors creates "high-risk" bookings. Our custom Cancellation Risk Score (developed in 04_Advanced_Analysis.ipynb) shows that bookings with long lead times (>90 days), made via an OTA for a 5-star hotel, have a cancellation rate of over 60%. This confirms that risk is predictable and concentrated in specific segments.

Why do some channels perform better?

- **Root Cause:** Differing customer intent. Direct and Corporate channels have the lowest cancellation rates because their users have a firm intention to stay. In contrast, OTA users often exhibit "shopping cart" behavior, booking multiple options and cancelling later. While OTAs drive volume, direct channels deliver more profitable and reliable revenue.

4. Strategic Recommendations

Based on the analysis, we propose the following data-driven strategies.

Strategy 1: Proactively Manage Cancellations Using the Risk Score

Goal: To reduce the overall cancellation rate by targeting high-risk bookings.

Recommendation 1.1: Implement Tiered Booking Policies.

- **Action:** For bookings with a CancellationRiskScore of 2 or 3, require a non-refundable deposit or shorten the free cancellation window.
- **Impact:** Deters speculative bookings and secures revenue from high-risk segments.

Recommendation 1.2: Launch a "Confirm & Commit" Campaign.

- **Action:** For high-risk bookings, send a targeted email 60 days before check-in offering a small incentive (e.g., free breakfast) to convert the booking to a non-refundable rate.
- **Impact:** Turns a high-risk booking into a committed one, improving revenue forecasting.

Strategy 2: Maximize Profitability by Driving Direct, High-Value Bookings

Goal: To shift the booking mix towards more profitable channels and customer segments.

Recommendation 2.1: Target "High-Value Guest" Profiles in Marketing.

- **Action:** Use the data-defined High-Value Guest Profile (detailed in 04_Advanced_Analysis.ipynb—customers preferring 4-star hotels with longer stays) to create targeted digital marketing campaigns.
- **Impact:** Attracts the most profitable customer segment, increasing average revenue per booking.

Recommendation 2.2: Enhance the Direct Booking Value Proposition.

- **Action:** Guarantee the "Best Rate" on the direct website and offer exclusive perks not available on OTAs (e.g., late checkout priority).
- **Impact:** Provides a compelling reason for customers to book direct, reducing commission costs.

Strategy 3: Optimize Revenue with a Data-Driven Seasonal Strategy

Goal: To maximize revenue during peak seasons and stimulate demand during off-seasons.

Recommendation 3.1: Implement Aggressive Peak-Season Pricing.

- **Action:** Confidently increase ADR by 15-20% during June-August and December, leveraging the clear seasonal demand.
- **Impact:** Directly increases high-season revenue.

Recommendation 3.2: Create Off-Season "Value Packages".

- **Action:** During low-demand months (e.g., Jan, Feb, Nov), create bundled packages to attract price-sensitive travelers and increase occupancy.
- **Impact:** Generates incremental revenue during otherwise slow periods.

5. Conclusion

This analysis has provided a clear, data-driven path to improving business performance. By implementing these strategies, the company can expect to see a measurable reduction in cancellations, an increase in direct, high-margin bookings, and a more profitable seasonal revenue model. The Jupyter notebooks provided contain all the code and visualizations that support these findings.

Appendix

```
--- Data Overview ---
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30000 entries, 0 to 29999
Data columns (total 24 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   customer_id        30000 non-null  int64  
 1   property_id        30000 non-null  int64  
 2   city               30000 non-null  object 
 3   star_rating        30000 non-null  int64  
 4   booking_date       30000 non-null  object 
 5   check_in_date      24532 non-null  object 
 6   check_out_date     24532 non-null  object 
 7   room_type          30000 non-null  object 
 8   num_rooms_booked   30000 non-null  int64  
 9   stay_type          30000 non-null  object 
 10  booking_channel    30000 non-null  object 
 11  booking_value      30000 non-null  float64
 12  costprice          30000 non-null  int64  
 13  markup             30000 non-null  int64  
 14  selling_price      30000 non-null  int64  
 15  payment_method     30000 non-null  object 
 16  refund_status      30000 non-null  object 
 17  refund_amount      30000 non-null  float64
 18  channel_of_booking 30000 non-null  object 
 19  booking_status     30000 non-null  object 
 20  travel_date        30000 non-null  object 
 21  cashback           30000 non-null  float64
 22  coupon_redeem      30000 non-null  float64
 23  Coupon USed?       30000 non-null  object 
dtypes: float64(4), int64(7), object(13)
memory usage: 5.5+ MB
```

Figure 8: Description of data

```
Feature engineering complete. Preview of new columns:
```

	booking_date	length_of_stay	is_cancelled	avg_daily_rate
0	2024-04-01	2	0	9680.495160
1	2024-04-01	7	0	876.714160
2	2024-04-01	5	0	4540.399092
3	2024-04-01	6	0	5677.999054
5	2024-04-01	3	0	1483.999505

Figure 9: Feature Engineering

```
Cleaned dataset 'hotel_bookings_cleaned.csv' loaded.
```

	customer_id	property_id	city	star_rating	booking_date	check_in_date	check_out_date	room_type	num_rooms_booked	stay_type	...	travel_date	cashback
0	492	3	San Francisco	4	2024-04-01	2024-05-24	2024-05-26	Standard	1	Leisure	...	2024-03-04	5.374694
1	180	3	Dallas	3	2024-04-01	2024-05-10	2024-05-17	Deluxe	1	Leisure	...	2024-07-19	7.161033
2	50	5	Dallas	3	2024-04-01	2024-05-31	2024-06-05	Deluxe	1	Business	...	2024-03-22	0.000000
3	294	3	Orlando	4	2024-04-01	2024-04-18	2024-04-24	Deluxe	3	Leisure	...	2024-11-24	7.932170
4	50	3	New York	2	2024-04-01	2024-04-18	2024-04-21	Standard	1	Leisure	...	2024-12-22	8.519053

5 rows × 30 columns

Figure 10: Cleaned Dataset